```
1
2
3   from a0_items import *
4   from plg_Regex import *
5
6
7   def get_ALL_urls (url):
8
9      reqs = requests.get(url)
10     soup = BeautifulSoup(reqs.text, 'html.parser')
11
12     LIST_url = []
13
14     for link in soup.find_all('a'):
15
16        #print(link.get('href'))
17        LIST_url.append (link.get('href'))
18
19     return LIST_url
20
21
22  # print (get_ALL_urls (url= 'https://www.geeksforgeeks.org/'))
23
24
25
26  def get_domain_from_url (url):
27
28     t = urlparse(url).netloc
29
30     domain = ('.'.join(t.split('.')[-2:]))
31
32     return domain
33
34
35  '''
36  LIST_domain = []
37  for x in get_ALL_urls (url='https://journalists.feedspot.com/usa_news_websites/'):
38
39      LIST_domain.append (get_domain_from_url (url=x))
40      #print (get_domain_from_url (url=x))
41
42
43  LIST_domain = list(set(LIST_domain))
44
45
46  for y in LIST_domain:
47
48     print (y)
49
50  '''
51
52
53  def get_HTML_elements (url):
54
55     page = requests.get(url)
56     soup = BeautifulSoup(page.text, 'html.parser')
57
58     for ID in soup.find_all('div', id=True):
59        print(ID.get('id'))
60
61
62  #get_HTML_elements (url = 'https://www.google.com')
63
64
65     # target url
66  url = 'https://www.google.com'
67
```

```
'''
page = requests.get(url)
soup = BeautifulSoup(page.text, 'html.parser')

classes = []
for element in soup.find_all(title=True):
    classes.extend(element["title"])

print (classes)
'''

page = requests.get(url)
soup = BeautifulSoup(page.text, 'html.parser').prettify()
'''
for tag in soup.findAll():
    try:
        print(tag['title'])
    except KeyError:
        pass
'''

#print (soup)


def test_code ():
   for line in soup.splitlines():

      class_only = remove_all_before_and_after_symbols (symbol_1 = 'class=' , symbol_2=
      '>', string=line)

      #print (class_only)
      print (line)


'''
How to get Domain search query url:
1. Get all elements from Webpage (Home page)
2. Get text field input and search input elements only (we need search textbox only) -
Use id, title, class, name to check if its a text field input
3. Use id, title, class..etc to check if the input text field exists -->
https://pythonexamples.org/python-selenium-check-if-input-text-field-exists/
4. Enter 3 values into the text field input and click enter. you should be redirected to
a search result page
5. The search result page should have values that you input into
6. Get the full url of the search result page
7. Remove anything AFTER the search query, including the search query to get the Domain
search query url
8. Use more search query with the DSQU to ensure the DSQU is correct (correct means that
it directs to the correct search result page, with correct search results in html )
'''

```